

Riesz Logic

Daoud Clarke

Abstract

We introduce Riesz Logic, whose models are abelian lattice ordered groups, which generalise Riesz spaces (vector lattices), and show soundness and completeness. Our motivation is to provide a logic for distributional semantics of natural language, where words are typically represented as elements of a vector space whose dimensions correspond to contexts in which words may occur. This basis provides a lattice ordering on the space, and this ordering may be interpreted as “distributional entailment”. Several axioms of Riesz Logic are familiar from Basic Fuzzy Logic, and we show how the models of these two logics may be related; Riesz Logic may thus be considered a new fuzzy logic. In addition to applications in natural language processing, there is potential for applying the theory to neuro-fuzzy systems.¹

Index Terms

Vector Lattice, Riesz Space, Fuzzy Logic, Distributional Semantics

I. INTRODUCTION

Much of the original motivation for fuzzy logic revolved around linguistic intuitions, for example the notion that “tall” is not a black and white concept, but that there are degrees of tallness. Indeed, one of the proposed applications for these ideas was in linguistics [1]. However, these ideas were never directly adopted by the linguistics or computational linguistics community. Instead, fuzziness has crept into natural language semantics research by the widespread adoption of “distributional semantics”, in which the meaning of words is determined by the contexts in which they occur. These techniques typically represent word meanings as vectors over these contexts, which capture fuzzy relationships between word meanings.

One question that is now being studied is how these vector based representations can be related to older, logical representations of meaning, in which a sentence would typically be translated into a logical form. The goal of this paper is to show that the vector spaces used in distributional semantics, Riesz Spaces, can be considered as models for a logic, which we call Riesz Logic (RL). Our hope is that this will lead to a confluence of distributional and logical semantics, opening up new areas of research and new methods of tackling problems in natural language processing.

RL has the following inference rules:

$$\frac{\phi, \phi \rightarrow \psi}{\psi} \quad (\text{MP}) \qquad \frac{\phi \rightarrow \psi}{\phi \vee \chi \rightarrow \psi \vee \chi} \quad (\text{RI})$$

¹Submitted to IEEE Transactions on Fuzzy Logic, Copyright 2014 IEEE

and axioms:

$$(\phi \rightarrow \psi) \rightarrow ((\psi \rightarrow \chi) \rightarrow (\phi \rightarrow \chi)) \quad (\text{R1a})$$

$$((\psi \rightarrow \chi) \rightarrow (\phi \rightarrow \chi)) \rightarrow (\phi \rightarrow \psi) \quad (\text{R1b})$$

$$\phi \rightarrow \phi \vee \psi \quad (\text{R2})$$

$$\phi \vee \psi \rightarrow \psi \vee \phi, \quad (\text{R3})$$

$$(\phi \vee \psi) \vee \psi \rightarrow \phi \vee \psi \quad (\text{R4})$$

$$0 \rightarrow (\phi \rightarrow \phi) \quad (\text{R5a})$$

$$(\phi \rightarrow \phi) \rightarrow 0 \quad (\text{R5b})$$

$$((\phi \rightarrow \psi) \vee 0 \rightarrow (\psi \rightarrow \phi) \vee 0) \rightarrow (\psi \rightarrow \phi) \quad (\text{R6a})$$

$$(\psi \rightarrow \phi) \rightarrow ((\phi \rightarrow \psi) \vee 0 \rightarrow (\psi \rightarrow \phi) \vee 0) \quad (\text{R6b})$$

In this paper we prove the soundness and completeness of this logic with respect to abelian lattice ordered groups, which generalise Riesz spaces, with formulas interpreted as asserting positivity. In doing this, we relate RL to the Logic of Equilibrium, known as BAL [2].

II. BACKGROUND

A. Distributional Semantics

Distributional semantics (see [3] for a comprehensive overview) is founded on the idea that the meaning of words can be determined by observing the contexts in which they occur. This idea has its origin in the work of Firth [4] and Harris [5], and the philosophy of Wittgenstein [6]. This idea has lead to techniques which analyse large text corpora to build word representations. For example, Figure 1 shows a sample of occurrences of the word “fruit” in the British National Corpus. Word representations built from such corpora are typically vectors describing the frequency with which the word occurs in different contexts. Depending on the application, the set of contexts which form the basis for the vector space will vary:

- Document identifiers: the context of a word is treated as the ID of the document in which it occurs.
- Other words: a word is considered to cooccur with another word if they are seen together within a window of a fixed number of words, or within the same sentence.
- Grammatical relations: sentences may be parsed to give dependency relations between words, and these relations treated as contexts.

Table I shows hypothetical occurrences of a few terms where document identifiers have been used as the contexts.

These raw frequency vectors are then typically processed in a variety of ways to reliably determine relationships between words. A typical system will employ one or more of the following techniques:

- Stopword removal or feature selection to identify contexts that provide useful contributions to the word’s meaning

end some medicine for her, but she will need fruit and milk, and some other special things that our own. Here we give you ideas for foliage, fruit and various festive trimmings that you can i part II). However, other strategies can bear fruit and are described under three sections which supper tomatoes, potato chips, dried fruit and cake. And they drank water out of tea-cu erent days, as the East Berliners queue for fruit and cheap stereos, a Turkish beggar sleeps i dening; and Pests -- how to control them on fruit and vegetables. Both are produced by the Hen me, "Silver Queen" is male so will never bear fruit At the opposite end of the prickliness sca lifted away Like an orange lifted from a fruit-bowl And darkness, blacker Than an oil- ed in your wreath. Christmas ribbon and wax fruit can be added for colour. Essentials are scis e you need to start developing your very own fruit collection KEEPING OUT THE COLD Need e ly with Jeyes fluid THE KITCHEN GARDEN FRUIT Cut out cankers on fruit trees, except tho wn and watered AUTUMN HUES Foliage and fruit enrich the autumn garden, whether glowing th - have forgotten the maxim: "tel arbre tel fruit ". If I were willing to unstitch the past of three children of Alfred Roger Ackerley, fruit importer of London, and his mistress, Janett rful didactic spirit, much that was to bear fruit in his years as a mature artist. Although thi e all made with natural vegetable, plant and fruit ingredients such as chamomile, kukai nut and ack in the soup. He re-visits the Copella fruit juice farm in Suffolk, the business he told rategic relationship" with Lotus, the first fruit of which is a mail gateway between Office and , choose your plants carefully to enjoy the fruit of your labour all year round. PLACES TO V and I love chips. Otherwise I'll nibble on fruit or something to convince myself that I'm eat tone and felt the softness and warmth of a fruit ripening against a wall? If she had she migh ol place to set. Calories per slice: 395 Fruit Scones with cinnamon Butter (makes 12) ought me water. Another monster gave me some fruit to eat. A few monsters lay against my body a ney fungus. Cut out diseased wood on most fruit trees VEGETABLES Continue winter diggin age and chafing. Remove old, unproductive fruit trees by cutting them down to shoulder heigh ITCHEN GARDEN FRUIT Cut out cankers on fruit trees, except those on peaches, plums and ch ps remain, then stir in the sugar and dried fruit. Using a round- ended knife, stir in the mil of a homeland, well others dream too, De fruit was forbidden an now yu can't chew, How ca onnoisseurs. We take a bite from an unusual fruit. We come away neither nourished nor ravished,

Fig. 1: Occurrences and some context of occurrences of the word *fruit* in the British National Corpus.

- Replacing frequency counts with derived statistics such as pointwise mutual information or TF-IDF (term frequency-inverse document frequency)
- Dimensionality reduction such as random projection or truncated singular value decomposition.

After these processing stages, instead of being integer frequency counts, word vectors are now typically real-valued, and may even contain negative values, for example if pointwise mutual information is used (although it is also common to set the negative components to zero). There is also typically still a preferred basis for the vector space, although the meaning of the basis vectors may no longer be tied to individual contexts, for example if dimensionality reduction has been performed. Instead, these dimensions are often considered to correspond to hidden "latent" aspects of meaning.

B. Distributional Generality

Distributional word vectors are often used in applications where it is enough to know how similar two words are in meaning. For this purpose, measures of distributional *similarity* are sufficient, for example, the cosine of the

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
banana	2	–	–	–	5	–	5	–
apple	4	3	4	6	3	–	–	–
orange	–	2	1	–	–	7	–	3
fruit	–	1	3	–	4	3	5	3
tree	–	–	5	–	–	5	–	–
computer	–	–	–	6	–	–	–	–

TABLE I: A table of hypothetical occurrences of words in a set of documents, d_1 to d_8 .

angle between two vectors is one measure that is often used. However, for applications such as information retrieval or question answering it is important to know whether it is likely that one word *entails* another. Recent research has investigated to what degree it is possible to determine this from word vectors [7]–[11]. One proposal supported by these experiments, known as distributional *generality* or distributional *inclusion*, is the idea that words with a more general meaning will occur in a wider range of contexts.

This idea was formalised in [12] in terms of the lattice ordering that is implicit in the vector space. Since the vector spaces used in these applications almost always have a preferred basis, it is possible to define a lattice ordering, where the meet and join operations are the component-wise minimum and maximum respectively. This makes the space a vector lattice, or Riesz space:

Definition 1 (Partially ordered vector space). A partially ordered vector space V is a real vector space together with a partial ordering \leq such that:

$$\text{if } u \leq v \text{ then } u + w \leq v + w$$

$$\text{if } u \leq v \text{ then } \alpha u \leq \alpha v$$

for all $u, v, w \in V$, and for all $\alpha \geq 0$. Such a partial ordering is called a **vector space order** on V . An element u of V satisfying $u \geq 0$ is called a **positive element**; the set of all positive elements of V is denoted V^+ . If \leq defines a lattice on V then the space is called a **vector lattice** or **Riesz space**.

The intuition behind the ordering \leq is that it describes a *distributional entailment*: assuming that \hat{x} and \hat{y} are distributional vectors of word frequencies for words x and y respectively, then $\hat{x} \leq \hat{y}$ means that y occurs at least as frequently as x in all contexts. Figure 2 gives an example of the meet operation for hypothetical word frequency vectors.

III. INTERPRETATIONS

Later we will prove the soundness and completeness of RL with respect to abelian lattice ordered groups, which generalise Riesz spaces, with vector addition and negation forming the group operations.

Definition 2 (Lattice Ordered Group). A partially ordered group is a tuple $\langle G, +, \leq \rangle$ such that $\langle G, + \rangle$ is a group,

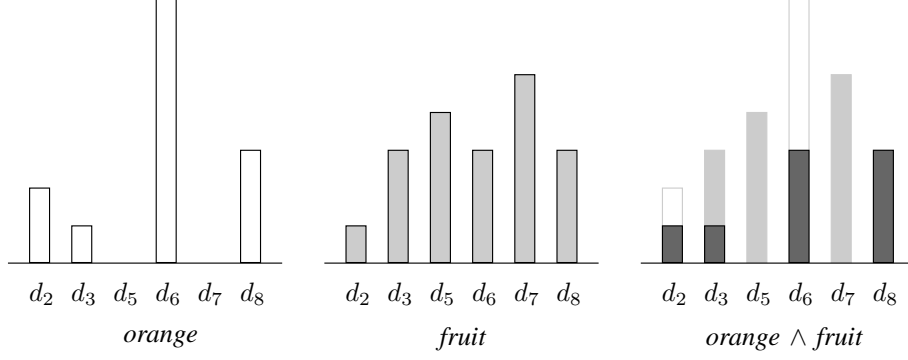


Fig. 2: Vector representations of the terms *orange* and *fruit* and their vector lattice meet (the darker shaded area).

and \leq is a partial order on G such that if $u \leq v$ then $u + w \leq v + w$ and $w + u \leq w + v$. If \leq is a lattice order, then G is called a lattice ordered group. Where there is no confusion, we refer to the lattice ordered group $\langle G, +, \leq \rangle$ as simply G . We denote the lattice meet and join by \wedge and \vee respectively.

The **positive part** of $u \in G$ is written u^+ and is defined as $u^+ = u \vee 0$; its **negative part** is defined as $u^- = (-u) \vee 0$.

Riesz spaces are abelian lattice ordered groups where the group operation is vector space addition, and the vector space zero is the unit of the group.

An interpretation $\langle G, F \rangle$ for RL is an abelian lattice ordered group G and a function F that maps variables in RL to elements of G . A formula x has the interpretation $\llbracket x \rrbracket$ defined recursively as follows:

- $\llbracket \phi \rrbracket = F(\phi)$
- $\llbracket x \rightarrow y \rrbracket = \llbracket y \rrbracket - \llbracket x \rrbracket$
- $\llbracket x \vee y \rrbracket = \llbracket x \rrbracket \vee \llbracket y \rrbracket$
- $\llbracket 0 \rrbracket = 0$

Note that the symbols \vee , \wedge and 0 are used both as symbols in the logic (on the left hand side) and in their vector space sense (on the right hand side).

The formula x is interpreted as asserting that $0 \leq \llbracket x \rrbracket$. Thus, for example, the formula $\phi \rightarrow \psi$ is interpreted as the assertion $0 \leq F(\psi) - F(\phi)$, or $F(\phi) \leq F(\psi)$. A formula x is satisfiable if there is some interpretation such that $0 \leq \llbracket x \rrbracket$; it is a theorem or tautology if $0 \leq \llbracket x \rrbracket$ for all interpretations.

IV. RELATION TO FUZZY LOGIC AND NEURAL NETWORKS

A. Fuzzy Logic

Our goal is to show that RL may be viewed as a type of fuzzy logic, although a non-standard one, both to aid in gaining an intuition for the nature of the logic, and to demonstrate its potential as a reasoning system. Firstly,

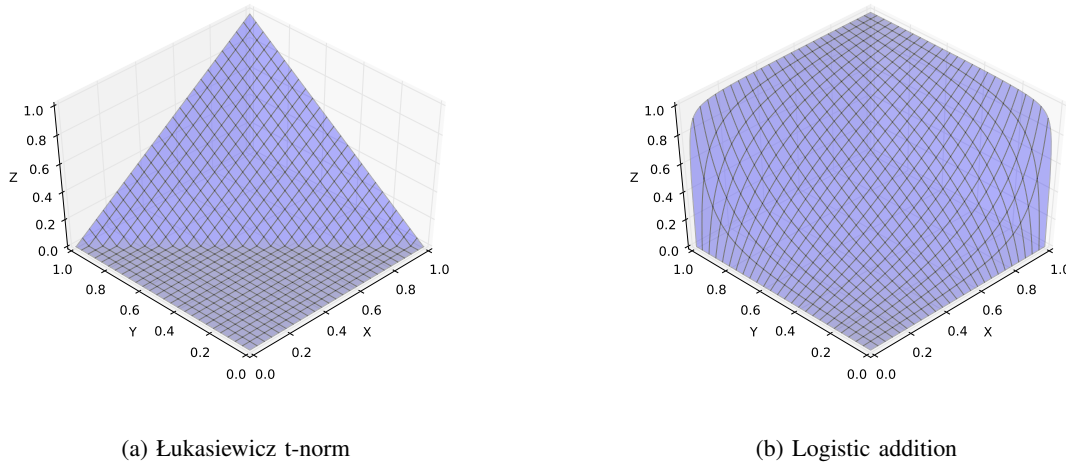


Fig. 3: Mapping real numbers to the interval $(0, 1)$ gives a new operation corresponding to addition (b), which bears some similarities to t-norms (a).

it is worth noting that there is some overlap between the axioms of RL and Basic Fuzzy Logic (BL): R1a is an axiom of BL, and R2–R4 hold in BL since \vee is a lattice join; other axioms are specific to RL.

Most fuzzy logics are interpreted in terms of the real interval $[0, 1]$. Consider the vector lattice of the real numbers (the single dimensional vector space). This can be mapped to the open interval $(0, 1)$, for example using the logistic function:

$$f(x) = \frac{1}{1 + e^{-x}}$$

The operations \wedge and \vee (maximum and minimum) behave the same when their behaviour is translated to this space.

Many fuzzy logics are derived from T-norms, which have the following properties:

$$\begin{aligned}
 T(a, b) &= T(b, a) && \text{(Commutativity)} \\
 T(a, b) &\leq T(c, d) \text{ if } a \leq c \text{ and } b \leq d && \text{(Monotonicity)} \\
 T(a, T(b, c)) &= T(T(a, b), c) && \text{(Associativity)} \\
 T(a, 1) &= a && \text{(Identity)}
 \end{aligned}$$

For example, the Łukasiewicz T-norm is defined as $T_L(a, b) = \max\{0, a + b - 1\}$.

A natural question to ask is whether we can define a T-norm for RL. Vector space addition seems like a natural candidate for this, because of its similarity to T_L . Note that in RL, addition \oplus can be defined by

$$\phi \oplus \psi := (\phi \rightarrow 0) \rightarrow \psi.$$

Addition of two real numbers translates to the interval $(0, 1)$ as:

$$T_R(a, b) = \frac{ab}{ab + (1 - a)(1 - b)}$$

See figure 3 for a three dimensional depiction of T_L and T_R . The first three of these properties are satisfied by T_R since they are properties of addition of real numbers, only the identity property is unsatisfied. Instead T_R is in general undefined for a or b equal to 1, although clearly for $a \neq 0$, $T_R(a, 1) = 1$, a quite different property from that of T-norms.

Thus RL has some very similar properties to fuzzy logics. The lattice operations of RL correspond to the weak conjunction and disjunction of fuzzy logics, whilst vector addition (defined implicitly via the \rightarrow operation) corresponds to strong conjunction. The major difference between RL and fuzzy logics is that there are no constants for “true” or “false”, only the constant 0 representing complete uncertainty.

B. Neural Networks

Recent work in deep neural networks [13] has shown that replacing the sigmoid function for activation with a “rectified linear unit” can improve accuracy and training time. A rectified linear unit is simply the function $f(x) = \max(x, 0)$, which is equivalent to $f(x) = x^+$ in vector lattice notation. This opens up the possibility of using RL in combination with neural networks, a potential new approach to neuro-fuzzy modelling [14]–[16].

V. SOUNDNESS

Proving soundness of the logic amounts to proving the validity of the rule and axioms.

MP: If $0 \leq F(\phi)$ and $0 \leq F(\psi) - F(\phi)$ then $0 \leq F(\psi)$ by transitivity of \leq . ■

RI: This follows from simple lattice-theoretic properties: The assertion $\phi \rightarrow \psi$ translates to $F(\phi) \leq F(\psi)$. As a shorthand, let us write $u = F(\phi)$, $v = F(\psi)$ and $w = F(\chi)$; we wish to show that $u \leq v$ implies $u \vee w \leq v \vee w$. To see this:

$$\begin{aligned} u &\leq v \\ u \vee v &= v \\ u \vee v \vee w &= v \vee w \\ (u \vee w) \vee (v \vee w) &= v \vee w \\ u \vee w &\leq v \vee w. \end{aligned}$$

RI: Since R1a is the converse of R1b, they may be taken together as asserting equality, by the antisymmetry of \leq . Thus we need to show:

$$\begin{aligned} \llbracket \phi \rightarrow \psi \rrbracket &= \llbracket (\psi \rightarrow \chi) \rightarrow (\phi \rightarrow \chi) \rrbracket \\ F(\psi) - F(\phi) &= \llbracket \phi \rightarrow \chi \rrbracket - \llbracket \psi \rightarrow \chi \rrbracket \\ &= F(\chi) - F(\phi) - F(\chi) + F(\psi) \\ &= F(\psi) - F(\phi). \end{aligned}$$

R2–4 are trivially seen to be properties of the partial ordering. R5 defines the symbol 0 such that $\llbracket 0 \rrbracket$ is the identity of the group, which we also denote 0. ■

R6:

$$\begin{aligned} \llbracket (\phi \rightarrow \psi) \vee 0 \rightarrow (\psi \rightarrow \phi) \vee 0 \rrbracket &= \llbracket \psi \rightarrow \phi \rrbracket \\ \llbracket (\psi \rightarrow \phi) \vee 0 \rrbracket - \llbracket (\phi \rightarrow \psi) \vee 0 \rrbracket &= F(\phi) - F(\psi) \\ (F(\phi) - F(\psi))^+ - (F(\phi) - F(\psi))^- &= F(\phi) - F(\psi) \end{aligned}$$

The identity $x = x^+ - x^-$ is well known for abelian lattice ordered groups, and can be shown by $x + x^- = x + (-x) \vee 0 = (x - x) \vee (x + 0) = 0 \vee x = x^+$. ■

VI. COMPLETENESS

We show completeness by relating RL to BAL [2]. The semantics of BAL is also abelian lattice ordered groups, but a statement is interpreted as stating equality with zero. BAL has the primitive binary operation \rightarrow and unary operation $^+$. The former is interpreted as in RL and the latter has the interpretation $\llbracket x^+ \rrbracket = \llbracket x \rrbracket^+$, i.e. it maps elements to their positive parts. Thus the statement $\phi \rightarrow \psi$ in BAL is interpreted as an assertion that $F(\psi) - F(\phi) = 0$ or $F(\phi) = F(\psi)$. The logic has the following axioms:

$$\begin{aligned} (\phi \rightarrow \psi) \rightarrow ((\chi \rightarrow \phi) \rightarrow (\chi \rightarrow \psi)) & \quad \text{(BALB)} \\ (\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow (\psi \rightarrow (\phi \rightarrow \chi)) & \quad \text{(BALC)} \\ ((\phi \rightarrow \psi) \rightarrow \psi) \rightarrow \phi & \quad \text{(BALN)} \\ \phi^{++} \rightarrow \phi^+ & \quad \text{(BALP)} \\ ((\psi \rightarrow \phi)^+ \rightarrow (\phi \rightarrow \psi)^+) \rightarrow (\phi \rightarrow \psi) & \quad \text{(BALO)} \end{aligned}$$

and the following inference rules:

$$\begin{array}{ll} \frac{\phi, \phi \rightarrow \psi}{\psi}, & \text{(BALMP)} \\ \frac{\phi}{\phi^+}, & \text{(BALPI)} \end{array} \quad \begin{array}{ll} \frac{\phi, \psi}{\phi \rightarrow \psi}, & \text{(BALG)} \\ \frac{(\phi \rightarrow \psi)^+}{(\phi^+ \rightarrow \psi^+)^+}. & \text{(BALMI)} \end{array}$$

Note that BAL has the same expressive power as RL: a statement x in the Logic of Equilibrium is equivalent to two statements, x and $x \rightarrow 0$ in Riesz Logic. Conversely, the statement x in Riesz Logic is equivalent to the statement $(x \rightarrow 0)^+$ in the Logic of Equilibrium: this is asserting that the negative part of x is zero, which is the same as asserting that x itself is positive.

Another consequence of the difference in interpretation between the two logics is that it is not enough to show that every tautology in BAL is a tautology in RL; we also expect their converses to hold. Our proof of completeness is thus in three parts:

- We show that the inference rules of BAL are valid in RL;
- We show that the axioms of BAL and their converses are tautologies of RL;
- We show that for every tautology in BAL of the form $(x \rightarrow 0)^+$, there is a tautology x in RL.

Proofs were constructed with the help of Prover9 [17].

A. Inference Rules

Premises in BAL inference rules are stronger statements than in RL since they are interpreted as asserting equality. Similarly, we need to deduce two conclusions in RL for each conclusion in a BAL inference rule in order to assert equality in RL. Specifically, given a BAL inference rule

$$\frac{\phi_1, \phi_2, \dots}{\psi},$$

we need the following inference rules in RL:

$$\frac{\phi_1, \phi_2, \dots, \phi_1 \rightarrow 0, \phi_2 \rightarrow 0, \dots}{\psi}$$

and

$$\frac{\phi_1, \phi_2, \dots, \phi_1 \rightarrow 0, \phi_2 \rightarrow 0, \dots}{\psi \rightarrow 0}.$$

For each rule BALR in BAL, we will refer to these two versions as BALR+ and BALR− respectively.

BALMP: BALMP+ follows trivially from the assumption of the rule MP in RL. To see BALMP−:

$$\alpha \rightarrow 0 \quad \text{(assumption)} \quad (1)$$

$$(\alpha \rightarrow \beta) \rightarrow 0 \quad \text{(assumption)} \quad (2)$$

$$(((\phi \rightarrow \psi) \rightarrow (\chi \rightarrow \psi)) \rightarrow \omega) \rightarrow ((\chi \rightarrow \phi) \rightarrow \omega) \quad \text{(MP, R1a, R1a)} \quad (3)$$

$$\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi) \quad \text{(MP, R1b, R1b)} \quad (4)$$

$$(0 \rightarrow \phi) \rightarrow ((\alpha \rightarrow \beta) \rightarrow \phi) \quad \text{(MP, 2, R1a)} \quad (5)$$

$$((\alpha \rightarrow 0) \rightarrow \phi) \rightarrow \phi \quad \text{(MP, 1, 4)} \quad (6)$$

$$(\phi \rightarrow \alpha) \rightarrow (\phi \rightarrow 0) \quad \text{(MP, 6, 3)} \quad (7)$$

$$(\alpha \rightarrow \beta) \rightarrow (\phi \rightarrow \phi) \quad \text{(MP, R5a, 5)} \quad (8)$$

$$\beta \rightarrow \alpha \quad \text{(MP, 8, R1b)} \quad (9)$$

$$\beta \rightarrow 0 \quad \text{(MP, 9, 7)} \quad (10)$$

■

BALPI: BALPI+ follows from R2. To see BALPI—:

(1)	$\alpha \rightarrow 0$	(assumption)	(1)
(2)	$\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi)$	(MP, R1b, R1b)	(2)
(3)	$\phi \vee \psi \rightarrow (\phi \vee \chi) \vee \psi$	(RI, R2)	(3)
(4)	$(\phi \vee \psi) \vee \chi \rightarrow (\psi \vee \phi) \vee \chi$	(RI, R3)	(4)
(5)	$(\phi \vee \psi \rightarrow \chi) \rightarrow (\psi \vee \phi \rightarrow \chi)$	(MP, R3, R1a)	(5)
(6)	$(\phi \vee \psi \rightarrow \chi) \rightarrow ((\phi \vee \psi) \vee \psi \rightarrow \chi)$	(MP, R4, R1a)	(6)
(7)	0	(MP, R3, R5b)	(7)
(8)	$((\phi \rightarrow \psi) \rightarrow \chi) \rightarrow (((\psi \rightarrow \phi) \vee 0 \rightarrow (\phi \rightarrow \psi) \vee 0) \rightarrow \chi)$	(MP, R6a, R1a)	(8)
(9)	$\alpha \vee \phi \rightarrow 0 \vee \phi$	(RI, 1)	(9)
(10)	$(0 \rightarrow \phi) \rightarrow \phi$	(MP, 7, 2)	(10)
(11)	$(0 \rightarrow \phi) \vee \psi \rightarrow \phi \vee \psi$	(RI, 10)	(11)
(12)	$(0 \vee \phi \rightarrow \psi) \rightarrow (\alpha \vee \phi \rightarrow \psi)$	(MP, 9, R1a)	(12)
(13)	$((\phi \vee \psi) \vee \chi \rightarrow \omega) \rightarrow (\phi \vee \chi \rightarrow \omega)$	(MP, 3, R1a)	(13)
(14)	$((\phi \vee \psi) \vee \chi \rightarrow \omega) \rightarrow ((\psi \vee \phi) \vee \chi \rightarrow \omega)$	(MP, 4, R1a)	(14)
(15)	$((\phi \rightarrow \psi \vee \chi) \vee 0 \rightarrow (\psi \vee \chi \rightarrow \phi) \vee 0) \rightarrow (\chi \vee \psi \rightarrow \phi)$	(MP, 5, 8)	(15)
(16)	$(\phi \vee \psi \rightarrow \chi) \rightarrow ((0 \rightarrow \phi) \vee \psi \rightarrow \chi)$	(MP, 11, R1a)	(16)
(17)	$(\phi \vee \psi) \vee \phi \rightarrow \psi \vee \phi$	(MP, R4, 14)	(17)
(18)	$\phi \vee \phi \rightarrow \psi \vee \phi$	(MP, 17, 13)	(18)
(19)	$\alpha \vee 0 \rightarrow \phi \vee 0$	(MP, 18, 12)	(19)
(20)	$(\alpha \vee 0) \vee 0 \rightarrow \phi \vee 0$	(MP, 19, 6)	(20)
(21)	$(0 \vee \alpha) \vee 0 \rightarrow \phi \vee 0$	(MP, 20, 14)	(21)
(22)	$(0 \rightarrow 0 \vee \alpha) \vee 0 \rightarrow \phi \vee 0$	(MP, 21, 16)	(22)
(23)	$\alpha \vee 0 \rightarrow 0$	(MP, 22, 15)	(23)
			(24)

■

BALG+:

- | | | |
|-----|---|----------------|
| (1) | $\alpha \rightarrow 0$ | (assumption) |
| (2) | β | (assumption) |
| (3) | $\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi)$ | (MP, R1b, R1b) |
| (4) | $((\phi \rightarrow \phi) \rightarrow \psi) \rightarrow (0 \rightarrow \psi)$ | (MP, R5a, R1a) |
| (5) | $(0 \rightarrow \phi) \rightarrow (\alpha \rightarrow \phi)$ | (MP, 1, R1a) |
| (6) | $(\beta \rightarrow \phi) \rightarrow \phi$ | (MP, 2, 3) |
| (7) | $0 \rightarrow \beta$ | (MP, 6, 4) |
| (8) | $\alpha \rightarrow \beta$ | (MP, 7, 5) |

■

Proof: BALG−

- | | | |
|-----|---|----------------|
| (1) | α | (assumption) |
| (2) | $\beta \rightarrow 0$ | (assumption) |
| (3) | $((\phi \rightarrow \psi) \rightarrow (\chi \rightarrow \psi)) \rightarrow \omega \rightarrow ((\chi \rightarrow \phi) \rightarrow \omega)$ | (MP, R1a, R1a) |
| (4) | $\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi)$ | (MP, R1b, R1b) |
| (5) | $((\beta \rightarrow 0) \rightarrow \phi) \rightarrow \phi$ | (MP, 2, 4) |
| (6) | $(\alpha \rightarrow \phi) \rightarrow \phi$ | (MP, 1, 4) |
| (7) | $(\phi \rightarrow \beta) \rightarrow (\phi \rightarrow 0)$ | (MP, 5, 3) |
| (8) | $(\alpha \rightarrow \beta) \rightarrow 0$ | (MP, 6, 7) |

■

BALMI: BALMI+ follows from R2. The proof of BALMI− is in three parts. The antecedent of BALMI− translates to the first assumption of the following proof, which demonstrates that if the positive part of $\alpha \rightarrow \beta$ is less than zero, then $\beta \rightarrow \alpha$:

- | | | |
|-----|--|---------------|
| (1) | $(\alpha \rightarrow \beta) \vee 0 \rightarrow 0$ | (assumption) |
| (2) | $(\phi \vee \psi \rightarrow \chi) \rightarrow (\phi \rightarrow \chi)$ | (MP, R2, R1a) |
| (3) | $(\alpha \rightarrow \beta) \rightarrow 0$ | (MP, 1, 2) |
| (4) | $(0 \rightarrow \phi) \rightarrow ((\alpha \rightarrow \beta) \rightarrow \phi)$ | (MP, 3, R1a) |
| (5) | $(\alpha \rightarrow \beta) \rightarrow (\phi \rightarrow \phi)$ | (MP, R5a, 4) |
| (6) | $\beta \rightarrow \alpha$ | (MP, 5, R1b) |

Given this, we then show that $(\alpha \vee 0 \rightarrow \beta \vee 0) \rightarrow 0$:

- (1) $\beta \rightarrow \alpha$ (assumption)
- (2) $((\phi \rightarrow \psi) \rightarrow (\chi \rightarrow \psi)) \rightarrow \omega \rightarrow ((\chi \rightarrow \phi) \rightarrow \omega)$ (MP, R1a, R1a)
- (3) $\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi)$ (MP, R1b, R1b)
- (4) $\beta \vee \phi \rightarrow \alpha \vee \phi$ (RI, 1)
- (5) $((\phi \rightarrow \phi) \rightarrow 0) \rightarrow \psi \rightarrow \psi$ (MP, R5b, 3)
- (6) $(\alpha \vee \phi \rightarrow \psi) \rightarrow (\beta \vee \phi \rightarrow \psi)$ (MP, 4, R1a)
- (7) $(\phi \rightarrow (\psi \rightarrow \psi)) \rightarrow (\phi \rightarrow 0)$ (MP, 5, 2)
- (8) $(\alpha \vee \phi \rightarrow \beta \vee \phi) \rightarrow 0$ (MP, 6, 7)

Finally, we make use of BALPI $-$ to show that we can take the disjunction with 0 on the left-hand side. Thus given $(\alpha \rightarrow \beta) \vee 0 \rightarrow 0$, we can show that $(\alpha \vee 0 \rightarrow \beta \vee 0) \vee 0 \rightarrow 0$. ■

B. Axioms

In this section, we show that the axioms of BAL are tautologies of RL. As before, we will refer to the axiom BALA of BAL as BALA+, and its converse as BALA $-$. Note that BALC is its own converse.

BALB+:

- (1) $((\phi \rightarrow \psi) \rightarrow (\chi \rightarrow \psi)) \rightarrow \omega \rightarrow ((\chi \rightarrow \phi) \rightarrow \omega)$ (MP, R1a, R1a)
- (2) $\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi)$ (MP, R1b, R1b)
- (3) $((\phi \rightarrow \psi) \rightarrow \psi) \rightarrow \chi \rightarrow (\phi \rightarrow \chi)$ (MP, 2, R1a)
- (4) $(\phi \rightarrow \psi) \rightarrow ((\chi \rightarrow \phi) \rightarrow (\chi \rightarrow \psi))$ (MP, 1, 3)

■

BALB-:

- (1) $\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi)$ (MP, R1b, R1b)
- (2) $((\phi \rightarrow \psi) \rightarrow \chi) \rightarrow (((\psi \rightarrow \omega) \rightarrow (\phi \rightarrow \omega)) \rightarrow \chi)$ (MP, R1b, R1a)
- (3) $((\phi \rightarrow \psi) \rightarrow (\chi \rightarrow \psi)) \rightarrow (((\chi \rightarrow \phi) \rightarrow \omega) \rightarrow \omega)$ (MP, 1, 2)
- (4) $((\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \chi)) \rightarrow (\psi \rightarrow \chi)$ (MP, 3, R1b)

■

BALC:

- (1) $((\phi \rightarrow \psi) \rightarrow (\chi \rightarrow \psi)) \rightarrow \omega \rightarrow ((\chi \rightarrow \phi) \rightarrow \omega)$ (MP, R1a, R1a)
- (2) $\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi)$ (MP, R1b, R1b)
- (3) $(\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow ((\omega \rightarrow \psi) \rightarrow (\phi \rightarrow (\omega \rightarrow \chi)))$ (MP, 1, 1)
- (4) $(\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow (\psi \rightarrow (\phi \rightarrow \chi))$ (MP, 2, 3)

■

BALN: BALN+ follows from Modus Ponens on R1a and R1b; BALN− follows from Modus Ponens applied to R1b twice. ■

BALP: BALP+ follows from R2; BALP− follows from R4. ■

BALO+ and BALO− are the only axioms we adopted unchanged in RL, as RL6a and RL6b respectively.

C. Equivalence of BAL and RL

For every tautology of the form $(\phi \rightarrow 0)^+$ in BAL, there is a tautology ϕ in RL.

Asserting Positivity:

- (1) $(\alpha \rightarrow 0) \vee 0 \rightarrow 0$ (assumption)
- (2) $((\phi \rightarrow \psi) \rightarrow \psi) \rightarrow \phi$ (MP, R1a, R1b)
- (3) $(\phi \vee \psi \rightarrow \chi) \rightarrow (\phi \rightarrow \chi)$ (MP, R2, R1a)
- (4) $(\alpha \rightarrow 0) \rightarrow 0$ (MP, 1, 3)
- (5) α (MP, 4, 2)

■

VII. CONCLUSION AND FUTURE WORK

We have described a new logic whose models are generalisations of vector lattices. Vector lattices are implicit in distributional representations of meaning used in many natural language processing applications, and our goal is to use the new logic to combine logical approaches to semantics with these distributional approaches. In order to achieve this, it is likely that several enhancements will need to be made to the logic, for example, we would probably need a first-order and perhaps higher-order versions to accurately represent natural language semantics.

It would also be interesting to extend the logic so that it is complete with respect to vector lattices; to do this, we would need some notion of multiplication by scalars.

ACKNOWLEDGMENTS

This work was funded by UK EPSRC project EP/IO37458/1 “A Unified Model of Compositional and Distributional Compositional Semantics: Theory and Applications”.

REFERENCES

- [1] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 1, pp. 28–44, 1973.
- [2] A. Galli, R. A. Lewin, and M. Sagastume, "The logic of equilibrium and abelian lattice ordered groups," *Archive for Mathematical Logic*, vol. 43, no. 2, pp. 141–158, 2004.
- [3] P. D. Turney and P. Pantel, "From frequency to meaning: Vector space models of semantics," *Journal of artificial intelligence research*, vol. 37, no. 1, pp. 141–188, 2010.
- [4] J. R. Firth, "A synopsis of linguistic theory, 1930–1955," in *Selected papers of JR Firth, 1952–59*. Indiana University Press, 1968, pp. 168–205.
- [5] Z. Harris, *Mathematical Structures of Language*. Wiley, New York, 1968.
- [6] L. Wittgenstein, *Philosophical Investigations*. New York: Macmillan, 1953, g. Anscombe, translator.
- [7] J. Weeds, D. Weir, and D. McCarthy, "Characterising measures of lexical distributional similarity," in *Proceedings of Coling 2004*. Geneva, Switzerland: COLING, Aug 23–Aug 27 2004, pp. 1015–1021.
- [8] M. Geffet and I. Dagan, "Lexical entailment and the distributional inclusion hypothesis," in *Proceedings of the 43rd meeting of the Association for Computational Linguistics (ACL)*, 2005, pp. 107–114.
- [9] I. Szpektor and I. Dagan, "Learning entailment rules for unary templates," in *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*. Manchester, UK: Coling 2008 Organizing Committee, August 2008, pp. 849–856. [Online]. Available: <http://www.aclweb.org/anthology/C08-1107>
- [10] L. Kotlerman, I. Dagan, I. Szpektor, and M. Zhitomirsky-Geffet, "Directional distributional similarity for lexical inference," *Special Issue of Natural Language Engineering on Distributional Lexical Semantics*, vol. 4(16), pp. 359–389, 2010.
- [11] J. Weeds, D. Clarke, J. Reffin, D. Weir, and B. Keller, "Learning to distinguish hypernyms and co-hyponyms," in *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014) August 2014*, 2014.
- [12] D. Clarke, "A context-theoretic framework for compositionality in distributional semantics," *Computational Linguistics*, vol. 38, no. 1, pp. 41–71, 2012.
- [13] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for lvcxr using rectified linear units and dropout," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 8609–8613.
- [14] C.-T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *Computers, IEEE Transactions on*, vol. 40, no. 12, pp. 1320–1336, 1991.
- [15] J.-S. Jang and C.-T. Sun, "Neuro-fuzzy modeling and control," *Proceedings of the IEEE*, vol. 83, no. 3, pp. 378–406, 1995.
- [16] S. Abe, *Pattern classification: neuro-fuzzy methods and their comparison*. Springer Publishing Company, Incorporated, 2012.
- [17] W. McCune, "Prover9 and mace4," 2005–2010, <http://www.cs.unm.edu/~mccune/prover9/>.